Supplementary material:

Genetic and Environmental Influences on Individual Differences in Sleep Duration During Adolescence

Saskia J. te Velde, Niels van der Aa, Dorret I. Boomsma, Eus J. W. van Someren, Eco J. C. de Geus, Johannes Brug, and Meike Bartels

A liability threshold model was used to analyze the ordinal data on sleep duration. The basic assumption underlying this model is that the effects of many genes and environmental factors add up to a continuous score that is assumed to be standard normal distributed (Falconer & Mackay, 1996). Ordinal scores on sleep duration were coded 0, 1, and 2. To model the three categories of sleep duration two thresholds were required. The thresholds, expressed in *z* values, are defined by the prevalence of the three categories of sleep duration in the sample and represent the value in the latent liability distribution above which an individual will endorse the next category.

Resemblance in sleep duration for twin and twin-sibling pairs on the liability scale is expressed in polychoric correlations. In a so-called saturated model, thresholds and polychoric twin and twin-sibling correlation were estimated for each of the five sex by zygosity groups (i.e., MZM, DZM, MZF, DZF, and DOS), making use of the software package Mx (Neale et al, 2006). The saturated model specifies for each sex by zygosity group that the data from the first- and second-born twin and the non-twin sibling are correlated without attempting to model these correlations as a function of genes or shared environment. In the saturated model, thresholds were estimated conditional on sex and age in order to assess sex and age differences in prevalence of sleep duration. Age, which was transformed into a *z* score, was included as a fixed effect (covariate) on both thresholds and it was allowed to differ for boys

and girls. Within a series of nested models we tested whether constraining the regression weights of age to be equal between boys and girls, constraining the regression weights of age at zero, and constraining the thresholds to be equal for boys and girls led to a significant deterioration of model fit. In addition, we tested whether twin and twin-sibling correlations were significantly different for MZ and DZ twin-sibling pairs.

Next, using structural equation modeling in Mx, genetic models were fitted to the data to estimate the influence of additive genetic (A), common environment shared by family members (C), and non-shared environment (E) to variation in sleep duration. To assess whether there was evidence for changes in the genetic architecture of sleep duration throughout adolescence, a moderator model was used (Purcell, 2002). Under this model, genetic, shared environmental, and non-shared environmental effects on the variance in a trait under study are modeled as a function of a moderator (i.e., age). However, under a liability threshold model, when the thresholds are estimated, the measurement scale of liability is unknown and assumed to be standard normal distributed with variance equal to 1. Thus, variance explained by A, C, and E sums to 1 and only two parameters need to be estimated. Here, what is of interest is whether the absolute magnitude of genetic, shared environmental, and non-shared environmental effects on the variance in a trait under study changes as a function of the moderator (Purcell, 2002). In order to be able to estimate the variance of the liability distribution, the two thresholds were constrained at 0 and 1 in the genetic models (in contrast to the saturated model in which thresholds were estimated). It has been demonstrated that this constraint enables estimation of the mean and variance of a latent liability distribution on an arbitrary scale, based on observed ordinal level data (Medland et al., 2009).

A graphical representation of the genetic model is given in Figure 1. The effect of age (*z* score) was included as a moderator on the genetic and environmental path coefficients *a*, *c*, and *e*. The influence of A, C, and E is represented by the paths $a + \alpha_{age}(\text{age})$, $c + \gamma_{age}(\text{age})$,

and $e + \eta_{age}(\text{age})$ and the variance due to A, C, and E was computed by squaring these paths. In order to assess sex differences in the genetic, shared environmental, and non-shared environmental contribution to variation in sleep duration, the α, γ, and η coefficients as well as the unmoderated path coefficients *a*, *c*, and *e* were allowed to be different for boys and girls, and it was tested whether constraining these parameter estimates to be equal for boys and girls resulted in a significant deterioration of model fit. The unmoderated estimates of *a*, *c*, and *e* represent the parameter estimates used to calculate the variance components A, C, and E at the mean age in the sample. The significance of the age effects was assessed by testing whether constraining the α, γ, and η coefficients at zero resulted in a significant deterioration of model fit.

Multiple models were fitted that were nested in the sense that one model could be derived from the other by the imposition of one or more constraints on the parameters. The fit of the different models was compared by means of the log-likelihood ratio test (LRT). The difference in minus two times the log-likelihood (-2LL) between two nested models has a χ^2 distribution with the degrees of freedom (*df*) equaling the difference in *df* between the two models. If a *p* value higher than .05 was obtained from the χ^2 test, the fit of the constrained model was not significantly worse than the fit of the more complex model. In this case, the constrained model was kept as the most parsimonious and best fitting model. The fit of the genetic models was also compared to the full ACE model by means of Akaike's Information Criterion (AIC), keeping the model with the lowest AIC as the best fitting model (Neale et al., 2006).

References

Falconer, D. S., & Mackay, T. F. C. (1996). *Introduction to quantitative genetics*. Essex, UK: Pearson Educational Limited.

Neale, M. C., Boker, S. M., Xie, G., & Maes, H. M. (2006). *Mx: Statistical modeling* (7th ed.). Richmond, VA: Department of Psychiatry.

Purcell, S. (2002). Variance components models for gene-environment interaction in twin analysis. *Twin Research, 5,* 554–571.

Medland, S. E., Neale, M. C., Eaves, L. J., & Neale, B. M. (2009). A note on the parameterization of Purcell's G × E model for ordinal and binary data. *Behavior Genetics, 39,* 220–229.